Introduction

The aim of this report is to provide insight into similar uses cases predominantly on Kaggle and other resources such as the journal outlined in the project plan. I will also conduct additional research into other areas apart from these two websites where similar use cases may have taken place.

Kaggle

Regarding searching for similar use cases on Kaggle there are no datasets or models that align directly with the aims of what we are trying to complete for the machine learning aspect of the project. The aim of our price prediction model is to use historical data of prices to predict future discounts using time-series models. The data available on Kaggle largely circles sales data, additionally the data we need is best sourced from the companies local to Australia.

This is true for the models available on Kaggle as well. There are a few dozen predictive models but none fall within the lines of price prediction for our use case.

Huggingface

Huggingface is a open-source platform for ML and data science. In similar fashion to Kaggle there were no similar use cases.

OpenML

OpenML provides a similar service to Kaggle and Huggingface. Providing a site which enables machine learning work to be shared. This includes datasets and machine learning models.

Overall, there weren't any similar use cases that could be found when searching through popular open-source machine learning and modelling sites. On Kaggle there were some resources which could be useful in further expanding the functionality of DiscountMate. Where work can be done with this is transaction data to enhance recommendations for users. Below is a small dataset but gives insight into what could be looked for to conduct modelling and data analysis using associative rules.

Transaction data

Below is another transactional dataset which was synthetically created that can help provide insight to enhance recommendations.

Transaction data 2

Price Prediction of Seasonal Items Using Time Series Analysis
(Ahmed Salah, 2023)

The purpose of this study was to look at price prediction of seasonal goods using machine learning based models and deep learning based models. Predictive models were used on a real-life dataset, using ensemble learning-based models, random forest, Ridge, Lasso, and Linear regression. It works on the belief that the study is the first work in the task of price prediction of seasonal items at its time of issue.

The dataset in this case consists of 18462 observations and contains information such as gift type, category, date of arrival in stock, date of stock update, buyer dates, whether the product is on sale or not, price and the quantity purchased. Features were also extracted from the date feature to represent time factor of the items. An 80-20 split was used for training and testing data and the Scikit-learn package was used.

The results were evaluated using 3 methods; residual histogram plots, true versus predicted values, and accuracy metric scores. The accuracy metric scores used in this study were Mean Absolute Error, Root Means Squared Error, R^2 and Mean Absolute Percentage Error.

Analysis of the histogram residuals show the use of ensemble learning models to produce the best performance followed by Random Forest, with the remaining ML and DL models providing similar performance. This same pattern is also demonstrated in the true versus predicted plot, however deep learning performed the worst in seasonal price prediction compared to other methods. Finally based off the accuracy metrics the top-performing models were the ensembles learning-based model, followed by random forest, DL models and remaining ML models respectively.

Insights

Based off the work completed in this article it indicates that the use of ensemble learning-based models provides better performance than other ML and DL models in predicting the price of seasonal items using time series analysis. While the dataset and aims don't directly fall in line within the scope of our project there are many attributes which can be used to benefit us. The use of ensemble learning-based approaches is something that can be studied and researched upon further.

Price Prediction of Seasonal Items Using Machine Learning and Statistical Methods
(Mohamed Ali Mohamed I. M.-h., 2022)

A similar study was also conducted that has close links with the time series analysis. With Ahmad Salah as an author on both. In this study a combination of four different ML models were used; support sector regressor, random forest, ridge and linear regression along with one statistical method, autoregressive integrated moving average. A Christmas gift dataset was used and the same metrics were used to evaluate the models.

Insights

Again, the random forest model produces the best outcomes, this time followed by the ARIMA model. The recommendation from the research is that employing the random forest model in conjunction with the ARIMA model to forecast seasonal product pricing. The authors also proposed addressing the seasonal items as a regression problem not a time series problem.

Conclusion

Overall, there has not been a significant amount of work done in modelling price prediction in the context Discount Mates is trying to work within. In a search of popular data science and modelling websites such as Kaggle and Huggingface no relevant resources were found. There were some resources which could be used to lead onto the addition of further features within the project, particularly surrounding recommendation systems and transactional data.

Finally, two relevant articles were found studying price prediction in seasonal items using modelling techniques. These did not directly link with the work being done in our project however some of the learning and concepts from these papers should be taken note of. Particularly the use of ensemble techniques to provide better predictive performance compared to other machine learning and deep learning techniques.